

Utah State University

**DigitalCommons@USU**

---

Instructional Technology and Learning Sciences  
Research

Instructional Technology & Learning Sciences

---

2012

## Multilinguality in the Digital Library: A Review

Anne Diekema

*Utah State University*

Follow this and additional works at: [https://digitalcommons.usu.edu/itls\\_research](https://digitalcommons.usu.edu/itls_research)



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Instructional Media Design Commons](#), and the [Teacher Education and Professional Development Commons](#)

---

### Recommended Citation

Diekema, A. R. (2012). Multilinguality in the digital library: A review. *The Electronic Library*, 30(2), 165-181.

This Article is brought to you for free and open access by the Instructional Technology & Learning Sciences at DigitalCommons@USU. It has been accepted for inclusion in Instructional Technology and Learning Sciences Research by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).





## Multilinguality in Digital Libraries: A Review

Journal:	<i>The Electronic Library</i>
Manuscript ID:	Draft
Manuscript Type:	Article
Keywords:	Digital libraries, Languages, Collaboration, Electronic libraries, Information retrieval

SCHOLARONE™  
Manuscripts

Review

Multilinguality in the Digital Library: A Review

1. Introduction

This article reviews literature on multilingual information access in digital libraries. As a natural consequence of increasing globalization and the advent and growth of the Internet, digital libraries have been created that not only cross borders, but also languages. The topic of the review covers the intersection of three research areas: cross-language information retrieval, multi-lingual information access, and digital libraries.

Cross-language information retrieval (CLIR) allows a user to query across languages (Peters and Sheridan, 2001, Oard and Diekema, 1998). Using CLIR in a collection containing documents with multiple languages, a source language query (e.g. in English) can retrieve relevant documents in one or more target languages (e.g., Dutch, Chinese, Arabic). The returned documents can then be translated into the source language for the English language reader. Multilingual information access (MLIA) is a broader term describing access (not limited to retrieval) to information across language boundaries and includes CLIR in addition to other areas such as multilingual summarization and cross-language question answering (Gey et al., 2006). A digital library (DL), sometimes referred to as electronic library or virtual library, is an organization that organizes and maintains an online collection of (born) digital or digitized materials and makes this collection available though the Internet (Lesk, 2005).

Multilingual digital libraries are digital libraries with content in more than one language, or that provide multilingual query access to a monolingual collection. The impetus behind the creation of multilingual digital libraries ranges from the practical to the grand and idealistic. Bringing together collections from various countries, regions, and cultures can provide quick and easy access to a wide range of information and ideas and can provide information access on a global scale (Yang et al., 2008, Maeda et al., 1998, Alessio, 2010), it also can preserve cultural heritage and advance agriculture (Jain and Gorla, 2006, Nichols et al., 2005). Multilingual digital libraries might also foster collaboration and lead to a deeper understanding between nations (Fox and Marchionini, 1998, Cousins, 2006).

Users of the multilingual digital libraries are quite diverse and range from children, economists, and cartographers, to European citizens (Bilal and Bachir, 2007a, Bilal and Bachir, 2007b, Geleijnse and Williams, 2007, Chias and Abad, 2009). Not surprisingly, the contents of these libraries reflect the diversity of their users. Among the multilingual content areas we found medical information, economics research, children's literature, Serbian culture, newspaper clippings, ancient Spanish maps, Indian theses, images, and legal literature (Lu et al., 2008, Zhou et al., 2006, Geleijnse and Williams, 2007, Hutchinson et al., 2005, Dudas, 2002, Calvanese et al., 2002, Tsang, 1997, Chias and Abad, 2009, Smits and Friis-Christensen, 2007, Clough and Sanderson, 2006, Urs et al., 2002, Francesconi and Peruginelli, 2004, Sheridan et al., 1997).

2. Scope

This review brings together writings on multilingual information access, or multilinguality, in digital libraries. Criteria for inclusion in the review are journal articles and journal news briefs (with authors) reporting on multilingual aspects in digital libraries. Selected articles had to be written in English, Dutch or German or be available in English translation and indexed in one or more of the databases described below. Articles on digital libraries or on cross-language information retrieval individually are not part of the review.

The literature included in this article resulted from searches in the following four electronic databases: ACM, ERIC, Library Literature, and Library, Information Science & Technology Abstracts. Search terms (keywords and subject or thesaurus terms) used were "digital libraries", "electronic libraries", "multilinguality", "multilingual", "multilingual materials", "cross-language information retrieval" and various Boolean combinations thereof. A few, additional articles were found in article bibliographies as well as serendipitously during the search process.

This is the first review on the topic of multilinguality in the digital library although reviews of related areas can be found. Oard and Diekema (1998) wrote an extensive review about research and practice in CLIR and, more recently, Kishida (2005) specifically focused on the technical issues that surface in CLIR. Bishop and Starr (1996) reviewed literature on social informatics of digital library use and infrastructure. Fox and Urs (2002) covered digital library research while Chowdhury (2006) examined articles on the usability of digital libraries and Bearman (2007) examined users and user needs. The most recent review we came across is on organizational and people issues in digital library research by Liew (2009).

### 3. Organization

The review begins with examples of existing multilingual DLs and the collaborative efforts and projects behind the creation of these types of libraries. Next, we discuss the aspect that differentiates the multilingual DL from the regular DL: crossing the language barrier and the resulting problems and challenges therein. The review concludes with research investigations and solutions for the cross-lingual problems follow.

### 4. Existing multilingual digital libraries

While many of the articles on multilingual DLs are made up of feasibility studies, prototype development efforts, announcements of future projects, or proposed frameworks (Lee et al., 2003, Braschler and Ferro, 2007, Amato et al., 2008, Chen and Ruiz, 2009, Jain and Gorla, 2006, Downing and Klein, 2001) a number of multilingual DLs are fully operational and examples from the literature are presented below. In their analysis of 150 digital libraries in the US, Chen and Ruiz (2009) found only five that were multilingual: Meeting of Frontiers, France in America, Parallel Histories, The International Children's Digital Library (ICDL), and The Perseus Digital Library. Out of these five only the ICDL appeared in the literature surveyed for this review.

The International Children's Digital Library (ICDL) [1] contains a collection of children's literature in over 50 languages [2]. It was created with funding from the National Science Foundation (NSF) and the Institute for Museum and Library Services (IMLS) as a collaborative project between the University of Maryland and the Internet Archive. The intended audiences of the ICDL are children, their parents, and other adult researchers interested in children's literature (Hutchinson et al., 2005). Interestingly, the project includes children as design partners and uses the ICDL as a platform for research on multilingual access (Hutchinson et al., 2005),(Bilal and Bachir, 2007a).

The World Digital Library (WDL) [3] contains primary source cultural content in a wide number of formats from a growing number of national libraries and other institutions and is operated by the United Nations Educational, Scientific and Cultural Organization (UNESCO) as well as the United States Library of Congress. The search interface is available in seven languages, ranked here in order of frequency of use: Spanish, English, Chinese, Portuguese, Russian, French and Arabic (Van Oudenaren, 2010). Sponsors of the library include Google and Microsoft. Van Oudenaren (2010) also reports over 10 million visitors since WDL's public launch. Most of these visitors hail from Argentina, Brazil, China, Columbia, France, Mexico, Portugal, Russia, Spain, and the United States.

Europeana [4]- the European Digital Library (TEL) - contains cultural and scientific materials and is funded by the European Union (Purday, 2009). The Europeans went through several stages of development that eventually resulted in Europeana. First, they created a website with information about national libraries of Europe as part of the Gabriel project (Gorman and Tran, 2002). Additional projects (TEL-ME-MOR) extended the national library collaboration to include national libraries from new European Union member countries and resulted in the integration of library resources and the ability to search across collections in the European Library (Cousins, 2006, Cousins et al., 2008, Braschler and Ferro, 2007, Fuegi and Segbert, 2006).

The digital library of the Caribbean (dLOC) [5] contains cultural, historical and research materials from the Caribbean (Wooldridge et al., 2009). DLOC is funded by a grant from the U.S. Department of Education and member contributions from dLOC partners.

Economists Online [6] contains references, open access full text publications and datasets by economists and is funded by the European Union through the Network of European Economists Online (NEEO) project (Geleijnse and Williams, 2007).

The Virtual Catalogue for Art History [7] contains bibliographic records from European art institutions. Formerly known as the Virtueller Katalog Kunstgeschichte (VKK) this meta catalog allows a search across the various catalogs. The project is funded by member fees (Hoyer, 2003).

Creating multilingual DLs requires collaboration on a large, often international, scale. Without exception the multilingual DLs described above are the result of collaborative projects. The next section describes some of the projects related to multilingual DLs that emerged from the literature.

## 5. Collaboration: Projects and Systems

While libraries have traditionally worked together in cooperative cataloging projects, museums with unique collections have often tended to work independently from each other (Kupietzky, 2007). In building a DL, collaboration is inevitable since it requires skills in various areas (e.g. computer science, library science, museum studies, arts, etc.) to build a digital library and additional skills such as linguistics, natural language processing, or cross-language information retrieval would be needed in cases of a multilingual DL. Funding from different agencies and foundations encourages collaboration in the form of projects with participants often hailing from several countries (Cousins et al., 2008, Gorman and Tran, 2002, Wooldridge et al., 2009) and between different organizations such as libraries and software companies (Calvanese et al., 2002). In some cases, these projects are collaboratively funded (Van Oudenaren, 2010). Other instances of collaboration manifest themselves in efforts to align national research agendas (Klavans and Schauble, 1998) or with the inclusion of proposed collaborative models in research papers (Lee et al., 2003, Cousins et al., 2008).

The majority of the research on multilingual DLs originates in Europe since cooperation between nations forms the foundation of the European Union and cross-lingual information navigation is imperative to Europe's daily operations. The European CACAO Project (cross-language access to catalogues and online libraries) incorporated cross-language information retrieval techniques into an infrastructure that allows users of online catalogs and digital libraries to query libraries in one European language and retrieve textual materials originating in other European languages. This multilingual navigational infrastructure has been incorporated in later European digital library projects such as the European Library (Levergood et al., 2008, Alessio, 2010). The project also participated in the European Cross-language Evaluation Forum CLEF (see section 8)(Bosca and Dini, 2009).

Another European project is the DELOS Network of Excellence. DELOS conducts research in the area of digital libraries, developing relevant technologies for all aspects of the library. One example of a technology is the DelosDLMS, which is a modular digital library management system with multilingual support (Braschler and Ferro, 2007, Ioannidis et al., 2008). Also European is the LAURIN project which is building a digital library of digitized multilingual newspaper clippings. Newspaper articles can be searched through the use of a multilingual thesaurus (Calvanese et al., 2002, Calvanese et al., 2001). We could not find an instance of LAURIN online.

The MultiMatch Project is creating a search engine for multilingual multimedia cultural heritage objects (Amato et al., 2007, Amato et al., 2008) while the Rastko project works to provide access to a multilingual Serbian culture collection (Dudas, 2002).

The systems described in the literature are typically created for research purposes. We only include systems here that have a direct application or connection to online multilingual collection searching.



MTIR is a Chinese-English information retrieval system that uses a bilingual dictionary for query translation in combination with a transliteration system to translated proper names. Multiple translation options are disambiguated using term co-occurrence information (Bian and Chen, 2000). Retrieved documents are translated using machine translation. Since the system is meant for use on the Web, the machine translations are carried out based on html tags. The system uses the HTTP protocol and can be easily integrated into web applications, potentially allowing for a bilingual online search. SPIRIT (Syntactic and Probabilistic Indexing and Retrieval of Information in Texts) originates from the 1980s as a monolingual, French and English system. It was expanded to be cross-lingual and uses reformulation rules that reformulate source queries into all possible target queries and use the document collection to disambiguate the translated queries (Fluhr et al., 1999). Eurovision is a cross-language image retrieval system that uses machine translation to translate queries into English which is then used to query the English-language index of the image captions (Clough and Sanderson, 2006). SIS-TMS is a thesaurus management system with a database scheme that allows storage and access to multiple multilingual thesauri and the connections between them (Doerr and Fundulaki, 1998). CLIR sometimes uses multilingual thesauri to get from the source language to the target language, which makes SIS-TMS especially relevant. Another terminological knowledge structure is used by SyDoM. SyDoM is a multilingual document system that uses a multilingual ontology to determine which terms to extract from documents for indexing (Roussey et al., 2007).

**6. Ways to cross the language barrier**

The unique feature about multilingual DLs is that they allow information searching across two or more different languages. Achieving this feat requires crossing the language barrier to match the information need (query) to content (documents) in disparate languages. According to the CLIR literature, there are many options to cross from one language to another (Yang and Li, 2005). One can translate the query into the language of the document, one can translate the documents into the language of the query, one can translate both query and document into a so-called interlingual representation. Interestingly, another translation approach has emerged in the multilingual DL literature: translation of metadata (or surrogate) records (Van Oudenaren, 2010, Lee et al., 2003, Francesconi and Peruginelli, 2004). Rather than translating an entire document it is much more efficient to translate the metadata record. This approach is especially suitable for collections with images and other non-textual materials that only have item level descriptions or metadata. A partial solution to crossing the language barrier is matching only on cognates (words that are shared between languages - typically proper names) in situations where the language scripts are the same (Buckley et al., 1998).

Translation knowledge comes from multilingual dictionaries, ontologies, and machine translation systems. This knowledge can be statistically extracted from text corpora. Most of these methods and translation knowledge sources are used in multilingual DLs.

Larson and cohorts (2002) harvest term translations from the library catalog of the University of California (over 10 million entries) to create a customized, multilingual

dictionary. Clinchant and Renders (2009) use a multi-language information retrieval model in combination with a multilingual dictionary that includes the source language.

Hodge (Hodge, 2000) provides an overview of knowledge organization systems (KOSs). Libraries have been organizing their collections using existing KOSs such as the Library of Congress Classification system or Library of Congress Subject Headings (LCSH) or their own homespun varieties. Unlike keywords, which are derived from bibliographic records or document full-text, subject terms are assigned to collection items by human catalogers or subject experts and provide high quality access points. Building on these knowledge systems in multilingual DLs requires a multilingual KOS such as a multilingual thesaurus. Schiel et al. use a computer supported indexing method to create a multilingual thesaurus (Schiel and Sousa, 2003, Schiel et al., 1999). Yang et al. (2008) create a multilingual thesaurus fully automatically by applying algorithms to a parallel aligned Chinese-English corpus. Calvanese et al. (2002) express the relationships between concepts in their multilingual thesaurus in a logical formalism used in query processing. The framework for the integration of multilingual heterogeneous thesauri is described by Nikolai et al. (1998). These thesauri can be used for indexing and browsed for retrieval. In the medical domain, Lu et al. (2008) developed a Chinese translation of the MeSH (Medical Subject Headings) in order to provide access to medical websites for Chinese language users. Smits and Friis-Christensen (2007) investigated whether it is desirable to use one common ontology which combines various structures together and concluded that creating such a structure is unrealistic. Sheridan et al. automatically created a similarity thesaurus drawn from a parallel corpus in the legal domain (Sheridan et al., 1997). While this structure is not a thesaurus in the strict sense of the word, the small groups of highly correlated multilingual terms function well to expand a monolingual query with multilingual terms.

Mixed translation methodologies are followed by Monroy et al. (2010) who use a multilingual glossary and an ontology, Brisaboa et al. (2002) use three dictionaries and ontologies to provide federated search capabilities across multiple digital libraries in different languages. Wang et al. (2006) mine the web to extend their translation dictionary.

## 7. Other problems and challenges

Crossing the language barrier is not the only challenge facing the multilingual DL. Additional problems and/or challenges are related to: data management and representation of information (Klavans and Schauble, 1998); interoperability (linking between systems) (Fox and Marchionini, 1998); development (Hutchinson et al., 2005); and copyright (which was not directly addressed in any of the retrieved articles).

### *Language barrier*

Cross-language information retrieval, which takes place in multilingual digital libraries, requires translation or crossing the language barrier. Errors introduced during the translation negatively impact search results in multilingual digital libraries. Even in a monolingual retrieval situation, lexical ambiguity and synonymy cause retrieval problems. With every additional language added to the mix, difficulties increase. Three factors cause translation errors: lack of translations for technical terms, acronyms, and proper names;



the erroneous breaking up of non-compositional phrases in translation, and; the addition of multiple translation senses of a word to the translation. In contrast to the CLIR literature, the multilingual digital library literature does not contain many articles about the translation problem. Problems with crossing the language barrier are most commonly associated with translation resources. Chung et al. and Wang et al. present a solution to the problem of missing dictionary terms in a query translation (Chung et al., 2004, Wang et al., 2006). Queries tend to be only a few words long and, when one or more words cannot be translated because they do not appear in the dictionary, retrieval is going to fail. The researchers mine the Web by extracting translations from bilingual search results and by adding these translations to the multilingual dictionary. Bian and Chen attempted to solve the missing dictionary terms by applying a transliteration algorithm to convert proper names, which make up a large percentage of untranslatable terms, from one writing system (Chinese) into another (English)(Bian and Chen, 2000).

*Data management*

Repository management and storage of content and metadata is a challenge to all digital libraries and has certain aspects that complicate matters for multilingual DLs. Instructions and metadata forms and vocabularies need to be translated into various languages and still make sense to users (Bia et al., 2005, Karvounarakis and Kapidakis, 2000). Digital library interfaces also need to be translated into the languages supported by the library, a process known as internationalization or localization. Multilingual document indexing is also challenging because each language has different characteristics and rules to decide between content bearing terms that you want to index and stop words that you want to remove (Roussey et al., 2007). Additional problems are created when using optical character recognition (OCR) on documents, especially for non-Latin character sets (Govindaraju et al., 2004).

*Representation*

Displaying text on a computer screen requires characters (letters or pictograms) to be represented as codes in the computer. This process is called character encoding and is required for readability, text processing such as indexing, and making text searchable. There is a confusing number of encoding schemes available but the most widely used encoding on the Internet is the UTF-8 Unicode encoding (Graham, 2000). According to the Unicode Consortium [8] Unicode can represent any character in any written language. Unfortunately not all existing languages are currently included. New languages are continually being added however. In 1998, Maeda et al. developed a technology to enable viewing of multilingual documents in a web browser (Maeda et al., 1998). Nowadays, these capabilities are included in most common browsers. That said, options might be limited for languages that are less common. Digital library software, such as Greenstone, support non-Latin character sets. This open-source software was used in a Mongolian newspaper project (Matusiak and Munkhmandakh, 2009). Since not all encoding schemes play nicely together and not all languages are represented in any of them, multilingual digital libraries may face a difficult challenge in finding a suitable encoding scheme.

*Development*

Most multilingual DL projects require (cross-cultural) collaboration (see section 5) and face the challenge of running projects successfully with large cross-cultural teams. Case studies are particularly well suited to learn more about the challenges involved in the development of multilingual DLs. Hutchinson et al. (2005) mention that the cultural aspects involved with building international software are perhaps the most complex issue to deal with. Cultural differences between different nationalities are often subtle and not easily understood by outsiders. They resolved this issue by creating a multicultural team to work on their library and advocate testing the interface with target audiences from all languages and cultures involved. A related problem these researchers reported is that of deciding what content was suitable (safe) for their underage users. Geleijnse and Williams (2007) reported another content development problem with the realization that they needed to get economists on board to contribute content to the DL. Liew (2005) warns us that creating a DL with indigenous cultural resources content requires careful representation as these resources can easily be misinterpreted. DL policies based on cultural norms of the cultural group are necessary. In her case study of the Asian Film Connection DL Afifi (2000) lists financial problems for DL development. It proved difficult to secure funding and resources for an educational DL project.

### *Interoperability*

Brisaboa et al. describe an architecture for combining digital libraries into one by creating a federated search interface (Brisaboa et al., 2002). The challenge here is to translate a single user query into the languages of all participating libraries, fire them off to each library and, finally, combine all the result sets into a single result list. Collaboration is intricately linked to interoperability. Operating in a multilingual, multicultural, and sometimes even multigenerational arena, is a challenge for the International Children's Library (Hutchinson et al., 2005). Similarly, the European Library project not only required interoperability on a technical level but also on socio-political and semantic levels (Cousins et al., 2008). Semantic interoperability sometimes occurs when combining different libraries with their different thesauri into one resulting in the intellectual and technical challenges of merging various knowledge structures (Kramer et al., 1997, McCulloch et al., 2005, Yang et al., 2008, Levergood et al., 2008).

## **8. Research on multilingual digital libraries**

Several articles mention or describe research agendas and research directions for multilingual information access (Klavans and Schauble, 1998, Gey et al., 2005, Gey et al., 2006). Common themes from these papers are: the introduction of real users or use cases in evaluation; extend the research to include more languages and media types; leverage experiences of real-world-deployments. Gey et al. (2005) add the goal of a multilingual web with search engines that search across languages and unified model for cross-language retrieval which combines the translation and retrieval components into a single algorithm. There is a significant body of research in CLIR but the uptake of this research into robust working systems with real users has thus far been limited (Gey et al., 2009, Peters, 2007).

System researchers typically build experimental systems to investigate certain approaches. They run experiments with their systems and compare results to decide whether certain approaches or algorithms are feasible and would lead to desired improvements. Much of the research in cross-language information retrieval takes place in various evaluation campaigns, in which participating systems are applied to various retrieval tasks. Systems are ranked in their abilities to complete these tasks successfully (Peters, 2007). While the TREC evaluation campaign included a cross-language evaluation track as early as 1997, the first evaluation campaign solely devoted to cross-language information retrieval is the NTCIR Workshop which started in 1999 (Kando et al., 1999) shortly followed by the Cross-Language Evaluation Forum (CLEF) which started in 2000 (Peters and Braschler, 2001). This long history of evaluation has resulted in a large collection of scientific data that can be used for further study. Agosti et al. propose to create a digital library to collect all of this data (Agosti et al., 2007). CLEF uses European languages and has developed tasks that are increasingly more realistic and relevant to the real world (Peters and Braschler, 2001, Peters, 2007). This is important since these evaluation campaign tasks challenge researchers and stimulate research in the task specific areas. Also, development teams of multilingual DLs are much more likely to participate in pragmatic evaluations because they require fewer system modifications and the results can be more easily applied. Examples of system-based research are described in the following section. The articles assembled here, with a few exceptions, have been selected based on their connection to multilingual digital libraries (a connection made either in the title or text of the article).

Query translation is a common approach to cross the language barrier and is highly applicable to multilingual DLs. Wang et al. describe a query translation system which can be connected to any digital library with monolingual (Chinese or English) content (Wang et al., 2004, Wang et al., 2006). This system mines the web for translations that do not occur in the dictionary (new terms, proper names). Although the researchers report "promising" results, it does not appear the system is ready for the real world just yet. Another query translation study was carried out by Bosca and Dini (2009) who participated in the CLEF evaluation forum with their system using various ways to expand the query with additional terms. The system performed well compared to other evaluation participants. In their CLEF experiments Clinchant and Renders (2009) used multi-language query translation trying to capitalize on multilingual documents (documents containing more than one language) but the results did not show improvement in retrieval results. Braschler and Ferro (2007) carried out a feasibility study to pick between two translation approaches (query or record) and to determine if the two alternatives might be combined. Kanazawa and his fellow researchers (2001) experimented with query translation techniques, and Yang et al. (2008) researched two different algorithms for automatic thesaurus construction benchmarking them against an earlier technique. Azzopardi et al. (2007) experimented with a model to generate simulated known-item queries for experimental systems that were comparable to real human queries which is useful for testing systems and modeling user query behavior.

Another way to do cross-language research is to prototype the system which you want to eventually build so that you can test whether the approach works on a smaller scale. A prototype, experimental system was applied to integrate various ontologies by Smits and Friis-Christensen (2007) who concluded that this approach was not viable. Larson et al.

(2002) used a prototype system to create a multilingual, conceptual mapping resource based on data mined from a large library catalog (described earlier in this paper). Bamman et al. (2010) tested a method for transferring structural information such as XML tags and chapter and section information) from source documents to target (translated) documents and achieved high accuracy. Ferber (1997) tested a system that automatically assigns subject terms based on the title of the document. This system used a set of documents with manually assigned subject headings to determine what descriptors to assign to the new documents. Results were found to be variable.

The majority of multilingual DL research appears to be system based. However, we did come across a few user-centered studies as well. Bilal and Bachir carried out two research studies with child users of the International Children's Digital Library (Bilal and Bachir, 2007a, Bilal and Bachir, 2007b). In the first study, the researchers tested interface design while the second study they observed the child subjects conducting searches and they also held group interviews to study the subjects' information seeking behavior. A qualitative study of the bilingual thesaurus-based interface called Searchling was carried out by Stafford et al. (2008). Fifteen users were asked to carry out three structured tasks to test the system as to test whether it aided in query formulation. Cousins (2006) studied the effect of a portal on usability. Clough and Sanderson (2006) ran user experiments with their Eurovision cross-language image retrieval system based on two search tasks.

Another type of multilingual DL research is the case study. Researchers studied the development of a multilingual library (Afifi, 2000, Liew, 2005) or a multilingual knowledge management system (O'Leary, 2008).

## 9. Conclusion

The articles reviewed in this paper show that there are a limited number of existing multilingual digital libraries but that their number is growing. Creating a multilingual digital library is typically a collaborative effort between different organizations and people with different areas of expertise. Enabling users to search across languages requires translation resources to cross the language barrier, which can be challenging depending on the language and resource availability. Additional challenges were found to be in the data management, representation (dealing with different fonts and character codes), development (creating international software, cross-cultural collaboration), and interoperability (system architecture and data sharing). Research in multilingual digital libraries was mostly system based involving experimental systems or system prototypes. A small number of user studies involved testing user interfaces and the study of information seeking behavior. Based on the literature however, it is mostly unclear who is using existing multilingual DLs, nor do we have much information about the extent of this use. There might be no uptake of research in systems with actual users but there is no research on the users in actual systems either. We can conclude that existing multilingual DLs are understudied and remain a bit of an enigma. Evaluation campaigns need to emphasize realistic evaluations that include working systems and actual users or use case scenarios. This might be the only way to ensure that CLIR research will find its way into existing multilingual DLs and that the library users find their way into the research literature.



Websites

[1] <http://childrenslibrary.org>  
[2] <http://en.childrenslibrary.org/about/fastfacts.shtml>  
[3] <http://wdl.org>  
[4] <http://europeana.eu>  
[5] <http://dloc.com>  
[6] <http://www.economistsonline.org>  
[7] <http://artlibraries.net>  
[8] <http://www.unicode.org>

References

AFIFI, M. (2000) Asian Film Connection: Developing a Scholarly Multilingual Digital Library - A Case Study. *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries*. Springer-Verlag.

AGOSTI, M., NUNZIO, G. M. D. & FERRO, N. (2007) The importance of scientific data curation for evaluation campaigns. *Proceedings of the 1st international conference on Digital libraries: research and development*. Pisa, Italy, Springer-Verlag.

ALESSIO, B. (2010) CACAO Project Overview. *D-Lib Magazine*, 16, 9-9.

AMATO, G., CIGARR, J., GONZALO, J., PETERS, C. & SAVINO, P. (2007) MultiMatch --- Multilingual/Multimedia Access to Cultural Heritage. *Proceedings of the 11th European conference on Research and Advanced Technology for Digital Libraries*. Budapest, Hungary, Springer-Verlag.

AMATO, G., DEBOLE, F., PETERS, C. & SAVINO, P. (2008) The MultiMatch Prototype: Multilingual/Multimedia Search for Cultural Heritage Objects. *Proceedings of the 12th European conference on Research and Advanced Technology for Digital Libraries*. Aarhus, Denmark, Springer-Verlag.

AZZOPARDI, L., DE RIJKE, M. & BALOG, K. (2007) Building simulated queries for known-item topics: an analysis using six european languages. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. Amsterdam, The Netherlands, ACM.

BAMMAN, D., BABEU, A. & CRANE, G. (2010) Transferring structural markup across translations using multilingual alignment and projection. *Proceedings of the 10th annual joint conference on Digital libraries*. Gold Coast, Queensland, Australia, ACM.

BEARMAN, D. (2007) Digital Libraries. *Annual Review of Information Science and Technology*, 41, 223-272.

BIA, A., MALONDA, J., G, J. & MEZ (2005) Automating multilingual metadata vocabularies. *Proceedings of the 2005 international conference on Dublin Core and metadata applications: vocabularies in practice*. Madrid, Spain, Dublin Core Metadata Initiative.

BIAN, G.-W. & CHEN, H.-H. (2000) Cross-Language Information Access to Multilingual Collections on the Internet. *Journal of the American Society for Information Science v. 51 no. 3 (February 2000) p. 281-96*.

- 1  
2  
3 BILAL, D. & BACHIR, I. (2007a) Children's interaction with cross-cultural and multilingual  
4 digital libraries. II. Information seeking, success, and affective experience.  
5 *Information Processing & Management*, 43, 65-80. Available  
6 <http://informationr.net/ir/11-1/paper240.html>  
7  
8 BILAL, D. & BACHIR, I. (2007b) Children's interaction with cross-cultural and multilingual  
9 digital libraries: I. Understanding interface design representations. *Information*  
10 *Processing & Management*, 43, 47-64.  
11  
12 BISHOP, A. P. & STARR, S. L. (1996) Social informatics of digital library use and  
13 infrastructure. *Annual Review of Information Science and Technology*, 31, 301-401.  
14  
15 BOSCA, A. & DINI, L. (2009) Query expansion via library classification system. *Proceedings*  
16 *of the 9th Cross-language evaluation forum conference on Evaluating systems for*  
17 *multilingual and multimodal information access*. Aarhus, Denmark, Springer-Verlag.  
18  
19 BRASCHLER, M. & FERRO, N. (2007) Adding multilingual information access to the European  
20 library. *Proceedings of the 1st international conference on Digital libraries:*  
21 *research and development*. Pisa, Italy, Springer-Verlag.  
22  
23 BRISABOA, N. R., JOS, PARAM, R., PENABAD, M. R., PLACES, N. S., RODR, F. J. & GUEZ  
24 (2002) Solving Language Problems in a Multilingual Digital Library Federation.  
25 *Proceedings of the First EurAsian Conference on Information and Communication*  
26 *Technology*. Springer-Verlag.  
27  
28 BUCKLEY, C., MITRA, M., WALZ, J. & CARDIE, C. (1998) Using Clustering and  
29 SuperConcepts within SMART: TREC 6. *The sixth text retrieval conference*.  
30 Gaithersburg, MD, National Institute of Standards and Technology.  
31  
32 CALVANESE, D., CATARCI, T., LENZERINI, M. & SANTUCCI, G. (2002) The multilingual  
33 thesaurus of LAURIN. *Proceedings of the 14th international conference on*  
34 *Software engineering and knowledge engineering*. Ischia, Italy, ACM.  
35  
36 CALVANESE, D., CATARCI, T. & SANTUCCI, G. (2001) LAURIN: A Distributed Digital  
37 Library of Newspaper Clippings. *World Wide Web*, 4, 5-20.  
38  
39 CHEN, J. & RUIZ, M. (2009) Towards an integrative approach to cross-language information  
40 access for digital libraries. *SIGIR 2009 Workshop: Information Access in a*  
41 *Multilingual World*. Boston, Massachusetts, USA, ACM.  
42  
43 CHIAS, P. & ABAD, T. (2009) Geolocating and Georeferencing: GIS Tools for Ancient Maps  
44 Visualisation. *Proceedings of the 2009 13th International Conference Information*  
45 *Visualisation*. IEEE Computer Society.  
46  
47 CHOWDHURY, S., LANDONI, M. & GIBB, F. (2006) Usability and impact of digital libraries:  
48 a review. *Online Information Review*, 30, 656-680.  
49  
50 CHUNG, W., ZHANG, Y., HUANG, Z., WANG, G., ONG, T.-H. & CHEN, H. (2004) Internet  
51 searching and browsing in a multilingual world: an experiment on the Chinese  
52 business intelligence portal (CBizPort). *J. Am. Soc. Inf. Sci. Technol.*, 55, 818-831.  
53  
54 CLINCHANT, S. & RENDERS, J.-M. (2009) Multi-language models and meta-dictionary  
55 adaptation for accessing multilingual digital libraries. *Proceedings of the 9th Cross-*  
56 *language evaluation forum conference on Evaluating systems for multilingual and*  
57 *multimodal information access*. Aarhus, Denmark, Springer-Verlag.  
58  
59 CLOUGH, P. & SANDERSON, M. (2006) User experiments with the Eurovision cross-  
60 language image retrieval system. *Journal of the American Society for Information*  
*Science and Technology*, 57, 697-708.



- COUSINS, J. (2006) The European Library - pushing the boundaries of usability. *Electronic Library*, 24, 434-444.
- COUSINS, J., CHAMBERS, S. & MEULEN, E. V. D. (2008) Uncovering cultural heritage through collaboration. *International Journal on Digital Libraries*, 9, 125-138.
- DOERR, M. & FUNDULAKI, I. (1998) SIS - TMS: A Thesaurus Management System for Distributed Digital Collections. *Proceedings of the Second European Conference on Research and Advanced Technology for Digital Libraries*. Springer-Verlag.
- DOWNING, A. & KLEIN, L. R. (2001) A multilingual virtual tour for international students. *College & Research Libraries News*, 62, 500.
- DUDAS, A. (2002) The Rastko project: the electronic library of Serbian culture (English). *Konyvtari Figyelo (Library Review)*.
- FERBER, R. (1997) Automated Indexing with Thesaurus Descriptors: A Co-occurrence Based Approach to Multilingual Retrieval. *Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries*. Springer-Verlag.
- FLUHR, C., SCHMIT, D., ANDRIEUX, C., ORTET, P., FR, D, BISSON, R. & COMBET, V. (1999) Crosslingual Interrogation of Multilingual Catalogs. *Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries*. Springer-Verlag.
- FOX, E. A. & MARCHIONINI, G. (1998) Toward a Worldwide Digital Library. *Communications of the ACM*.
- FOX, E. A. & URS, S. R. (2002) Digital libraries. *Annual Review of Information Science and Technology*, 36, 503-589.
- FRANCESCONI, E. & PERUGINELLI, G. (2004) Opening the legal literature portal to multilingual access. *Proceedings of the 2004 international conference on Dublin Core and metadata applications: metadata across languages and cultures*. Shanghai, China, Dublin Core Metadata Initiative.
- FUEGI, D. & SEGBERT, M. (2006) TEL-ME-MOR: A Milestone on the Way to the European Digital Library. *Alexandria*, 18, 55-62.
- GELEIJNSE, H. & WILLIAMS, P. (2007) NEEO and Economists Online. *SCONUL Focus*, 33-35.
- GEY, F., KARLGREN, J. & KANDO, N. (2009) Information access in a multilingual world: transitioning from research to real-world applications. *SIGIR Forum*, 43, 24-28.
- GEY, F. C., KANDO, N., LIN, C.-Y. & PETERS, C. (2006) New directions in multilingual information access. *SIGIR Forum*, 40, 31-39.
- GEY, F. C., KANDO, N. & PETERS, C. (2005) Cross-Language Information Retrieval: the way ahead. *Information Processing & Management*, 41, 415-431.
- GORMAN, G. E. & TRAN, L. A. (2002) Gabriel, come blow your horn! *New Library World*.
- GOVINDARAJU, V., KHEDEKAR, S., KOMPALLI, S., FAROOQ, F., SETLUR, S. & VEMULAPATI, R. (2004) Tools for Enabling Digital Access to Multi-Lingual Indic Documents. *Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL '04)*. IEEE Computer Society.
- GRAHAM, T. (2000) *Unicode : a primer*, Foster City, CA, M & T Books.
- HODGE, G. (2000) Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.
- HOYER, R. (2003) The Virtueller Katalog Kunstgeschichte as a tool for international cooperation. *Art Libraries Journal*.

- HUTCHINSON, H. B., ROSE, A., BEDERSON, B. B., WEEKS, A. C. & DRUIN, A. (2005) The International Children's Digital Library: A Case Study in Designing for a Multilingual, Multicultural, Multigenerational Audience. *Information Technology & Libraries*, 24, 4-12.
- IOANNIDIS, Y., MILANO, D., SCHEK, H.-J. & SCHULDT, H. (2008) DelosDLMS. *International Journal on Digital Libraries*, 9, 101-114.
- JAIN, S. P. & GORIA, S. (2006) Digital Library for Indian Farmers (DLIF) using open source software: A strategic planning. *IFLA Conference Proceedings*, 1-10.
- KANAZAWA, T., AIZAWA, A., TAKASU, A. & ADACHI, J. (2001) The Effects of the Relevance-Based Superimposition Model in Cross-Language Information Retrieval. *Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries*. Springer-Verlag.
- KANDO, N., KURIYAMA, K., NOZUE, T., EGUCHI, K., KATO, H., HIDAKA, S. & ADACHI, J. (1999) The NTCIR Workshop : the First Evaluation Workshop on Japanese Text Retrieval and Cross-Lingual Information Retrieval. *Fourth International Workshop on Information Retrieval with Asian Languages*. Taipei, Taiwan
- KARVOUNARAKIS, G. & KAPIDAKIS, S. (2000) Submission and repository management of digital libraries, using WWW. *Computer Networks*.
- KISHIDA, K. (2005) Technical issues of cross-language information retrieval: a review. *Information Processing & Management*, 41, 433-455.
- KLAVANS, J. L. & SCHAÜBLE, P. (1998) NSF-EU multilingual information access. *Commun. ACM*, 41, 69-70.
- KRAMER, R., NIKOLAI, R. & HABECK, C. (1997) Thesaurus federations: loosely integrated thesauri for document retrieval in networks based on Internet technologies. *International Journal on Digital Libraries*, 1, 122.
- KUPIETZKY, A. S. G. (2007) *Subject access to a multilingual museum database : a step-by-step approach to the digitization process*, Westport, Conn., Libraries Unlimited.
- LARSON, R. R., GEY, F. & CHEN, A. (2002) Harvesting translingual vocabulary mappings for multilingual digital libraries. *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*. Portland, Oregon, USA, ACM.
- LEE, W., SUGIMOTO, S., NAGAMORI, M., SAKAGUCHI, T. & TABATA, K. (2003) A subject gateway in multiple languages: a prototype development and lessons learned. *Proceedings of the 2003 international conference on Dublin Core and metadata applications: supporting communities of discourse and practice---metadata research & applications*. Seattle, Washington, Dublin Core Metadata Initiative.
- LESK, M. (2005) *Understanding digital libraries*, Amsterdam, Elsevier.
- LEVERGOOD, B., FARRENKOPF, S. & FRASNELLI, E. (2008) The specification of the language of the field and interoperability: cross-language access to catalogues and online libraries (CACAO). *Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*. Berlin, Germany, Dublin Core Metadata Initiative.
- LIEW, C. L. (2005) Cross-Cultural Design and Usability of a Digital Library Supporting Access to Maori Cultural Heritage Resources. IN THENG, Y.-L. & FOO, S. (Eds.) *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific*. Hershey, PA, Idea Group Pub.

- 1  
2  
3 LIEW, C. L. (2009) Digital Library Research 1997-2007: Organisational and People Issues.  
4 *Journal of Documentation*, 65, 245-266.  
5  
6 LU, W.-H., LIN, R. S., CHAN, Y.-C. & CHEN, K.-H. (2008) Using Web resources to construct  
7 multilingual medical thesaurus for cross-language medical information retrieval.  
8 *Decis. Support Syst.*, 45, 585-595.  
9  
10 MAEDA, A., DARTOIS, M., FUJITA, T., SAKAGUCHI, T., SUGIMOTO, S. & TABATA, K.  
11 (1998) Viewing multilingual documents on your local Web browser. *Communications of*  
12 *the ACM*, 41, 64-65.  
13  
14 MATUSIAK, K. K. & MUNKHMANDAKH, M. (2009) A Newspaper/Periodical Digitization  
15 Project in Mongolia: Creating a Digital Archive of Rare Mongolian Publications.  
16 *Serials Librarian*, 57, 118-127.  
17  
18 MCCULLOCH, E., SHIRI, A. & NICHOLSON, D. (2005) Challenges and issues in terminology  
19 mapping: a digital library perspective. *Electronic Library*, 23, 671-677.  
20  
21 MONROY, C., FURUTA, R. & CASTRO, F. (2010) Using an ontology and a multilingual glossary  
22 for enhancing the nautical archaeology digital library. *Proceedings of the 10th annual*  
23 *joint conference on Digital libraries*. Gold Coast, Queensland, Australia, ACM.  
24  
25 NICHOLS, D. M., WITTEN, I. H., KEEGAN, T. T., BAINBRIDGE, D. & DEWSNIP, M. (2005)  
26 Digital libraries and minority languages. *New Review of Hypermedia & Multimedia*,  
27 11, 139-155.  
28  
29 NIKOLAI, R., TRAUPE, A. & KRAMER, R. (1998) Thesaurus Federations: A Framework for  
30 the Flexible Integration of Heterogeneous, Autonomous Thesauri. *Proceedings of*  
31 *the Advances in Digital Libraries Conference*. IEEE Computer Society.  
32  
33 O'LEARY, D. E. (2008) A multilingual knowledge management system: A case study of FAO  
34 and WAICENT. *Decision Support Systems*, 45, 641-661.  
35  
36 OARD, D. W. & DIEKEMA, A. R. (1998) Cross-Language Information Retrieval. *Annual Review*  
37 *of Information Science and Technology*, 33, 223-256.  
38  
39 PETERS, C. (2007) Multilingual information access: the contribution of evaluation.  
40 *Proceedings of the 2006 international workshop on Research issues in digital*  
41 *libraries*. Kolkata, India, ACM.  
42  
43 PETERS, C. & BRASCHLER, M. (2001) European research letter: Cross-language system  
44 evaluation: The CLEF campaigns. *Journal of the American Society for Information*  
45 *Science and Technology*, 52, 1067-1072.  
46  
47 PETERS, C. & SHERIDAN, P. (2001) Multilingual access for information systems. *IFLA*  
48 *Conference Proceedings*, 1-8.  
49  
50 PURDAY, J. (2009) Think culture: Europeana.eu from concept to construction. *Electronic*  
51 *Library*, 27, 919-937.  
52  
53 ROUSSEY, C., CALABRETTO, S. & HARRATHI, F. (2007) Multilingual extraction of  
54 semantic indexes. *Proceedings of the 2007 international workshop on Semantically*  
55 *aware document processing and indexing*. Montpellier, France, ACM.  
56  
57 SCHIEL, U. & SOUSA, I. M. S. F. D. (2003) Interactive indexing of documents with a  
58 multilingual thesaurus. *Effective databases for text & document management*. IGI  
59 Publishing.  
60  
61 SCHIEL, U., SOUSA, I. M. S. F. D. & FERNEDA, E. (1999) SIM - A System for Semi-  
62 Automatic Indexing of Multilingual Documents. *Proceedings of the 10th*  
63 *International Workshop on Database & Expert Systems Applications*. IEEE  
64 Computer Society.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- SHERIDAN, P., BRASCHLER, M. & SCHAÜBLE, P. (1997) Cross-Language Information Retrieval in a Multilingual Legal Domain. *Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries*. Springer-Verlag.
- SMITS, P. C. & FRIIS-CHRISTENSEN, A. (2007) Resource Discovery in a European Spatial Data Infrastructure. *IEEE Trans. on Knowl. and Data Eng.*, 19, 85-95.
- STAFFORD, A., SHIRI, A., RUECKER, S., BOUCHARD, M., MEHTA, P., ANVIK, K. & ROSSELLO, X. (2008) Searchling: User-Centered Evaluation of a Visual Thesaurus-Enhanced Interface for Bilingual Digital Libraries. *Proceedings of the 12th European conference on Research and Advanced Technology for Digital Libraries*. Aarhus, Denmark, Springer-Verlag.
- TSANG, S. S. Y. (1997) Multilingual newspaper clippings image database (poster). *Proceedings of the second ACM international conference on Digital libraries*. Philadelphia, Pennsylvania, United States, ACM.
- URS, S. R., HARINARAYANA, N. S. & KUMBAR, M. (2002) A Multilingual Multi-script Database of Indian Theses: Implementation of Unicode at Vidyanidhi. *Proceedings of the 5th International Conference on Asian Digital Libraries: Digital Libraries: People, Knowledge, and Technology*. Springer-Verlag.
- VAN OUDENAREN, J. (2010) Connecting the World, Responding to User Needs. *Information Outlook*, 14, 10-12.
- WANG, J.-H., LU, W.-H. & CHIEN, L.-F. (2004) Toward Web mining of cross-language query translations in digital libraries. *International Journal on Digital Libraries*, 4, 247-257.
- WANG, J.-H., TENG, J.-W., LU, W.-H. & CHIEN, L.-F. (2006) Exploiting the Web as the multilingual corpus for unknown query translation. *Journal of the American Society for Information Science & Technology*, 57, 660-670.
- WOOLDRIDGE, B., TAYLOR, L. & SULLIVAN, M. (2009) Managing an Open Access, Multi-Institutional, International Digital Library: The Digital Library of the Caribbean. *Resource Sharing & Information Networks*, 20, 35-44.
- YANG, C. & LI, K. W. (2005) Cross-Lingual Information Retrieval: The Challenge in Multilingual Digital Libraries. IN THENG, Y.-L. & FOO, S. (Eds.) *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific*. Hershey, PA Idea Group Pub.
- YANG, C. C., WEI, C.-P. & LI, K. W. (2008) Cross-lingual thesaurus for multilingual knowledge management. *Decis. Support Syst.*, 45, 596-605.
- ZHOU, Y., QIN, J. & CHEN, H. (2006) CMedPort: an integrated approach to facilitating Chinese medical information seeking. *Decis. Support Syst.*, 42, 1431-1448.